The outline for Unit 2

UNIT 1. Introduction: The regression model. ✓

UNIT 2. Estimation principles.

2.1 Ordinary Least Squares.
2.2 Maximum Likelihood.
2.3 Method of Moments.

Addendum A. A brief revision of asymptotic theory.

Addendum B. Estimating the OLS estimator's distribution using bootstrap.

UNIT 3: Hypothesis testing principles.

UNIT 4: Heteroscedasticity in the regression model.

UNIT 5: Endogeneity of regressors.

The outline for today

Addendum A. A brief revision of asymptotic theory.

2.1 Ordinary Least Squares.

- Small samples properties.
- Asymptotic properties.

Recommended readings: Chapter 6 of Creel (2006) and Chapter 9 of Greene (2000).

To learn more: Chapters I to V of Halbert White (1984) *Asymptotic theory for econometrician*, Academic Press, Inc.

A brief revision of asymptotic theory

- Convergence in probability $\rightsquigarrow$ Weak consistency.
- Almost sure convergence $\rightsquigarrow$ Strong consistency.
- Complete convergence.
- Convergence in mean $\rightsquigarrow$ LLN.
- Convergence in distribution $\rightsquigarrow$ CLT.
- Stochastic orders.

Convergence in probability

Definition 1: Let $\{b_n(\omega)\}$ be a sequence of real-valued random variables. If there exists a real-valued random variable $b(\omega)$ such that for every $\varepsilon > 0$,

$$\Pr\{|b_n(\omega) - b(\omega)| < \varepsilon\} \to 1 \text{ as } n \to \infty,$$

then we say that $\{b_n(\omega)\}$ *converges in probability* to b.

Example 1: Let $\{b_n(\omega)\}$ a sequence s.t. $\Pr\{b_n(\omega) = 1\} = \frac{1}{n}$ and $\Pr\{b_n(\omega) = 0\} = 1 - \frac{1}{n}$, then

$$b_n \xrightarrow{p} 0.$$

Other frequent notation is $P \lim b_n = b$.

Almost sure convergence

Definition 2: Let $\{b_n(\omega)\}$ be a sequence of real-valued random variables. If there exists a real-valued random variable $b(\omega)$ such that

$$\Pr\{\omega : b_n(\omega) \to b(\omega)\} = 1,$$

then we say that $\{b_n(\omega)\}$ *converges almost surely* to b.

Proposition 1: Let $\{b_n(\omega)\}$ be a sequence of real-valued r.v. and $b(\omega)$ a real-valued r.v., $b_n \xrightarrow{a.s.} b$ *iff* for every $\varepsilon > 0$,

$$\lim_n \Pr\{\omega : \sup_{m \geq n} |b_m(\omega) - b(\omega)| > \varepsilon\} = 0.$$

Example 2: Let $\{b_n(\omega)\}$ a sequence s.t. $\Pr\{b_n(\omega) = 1/n\} = \frac{1}{2}$ and $\Pr\{b_n(\omega) = -1/n\} = \frac{1}{2}$, then

$$b_n \xrightarrow{a.s.} 0.$$

Almost sure and in probability convergence - 1

Proposition 2: Let $\{b_n(\omega)\}$ be a sequence of real-valued r.v. and $b(\omega)$ a real-valued r.v., such that $\{b_n(\omega)\}$ converges almost surely to b, then $\{b_n(\omega)\}$ converges in probability to b.

Proof: Without lost of generality we can assume that $\Leftarrow$ Warning

$b = 0$. So, if $b_n \xrightarrow{a.s.} 0$, for every $\varepsilon > 0$ and $\delta > 0$ we can choose an $n_0 = n_0(\varepsilon, \delta)$ s.t.

$$\Pr\left\{\bigcap_{n=n_0}^{\infty} |b_n| \leq \varepsilon\right\} \geq 1 - \delta,$$

then, for $n > n_0$ we have

$$\Pr\{|b_n| \leq \varepsilon\} \geq \Pr\left\{\bigcap_{n=n_0}^{\infty} |b_n| \leq \varepsilon\right\} \geq 1 - \delta,$$

and this implies that $b_n \xrightarrow{p} 0$.■

Almost sure and in probability convergence - 1

But the reciprocal isn't true.

Example 3: For every positive integer n we can find two integer m and k s.t. $n = 2^k + m$ and $0 \leq m < 2^k$. Let $\{b_n(\omega)\}$ a sequence defined by

$$b_n(\omega) = \begin{cases} 2^k & \text{if } m/2^k \leq \omega \leq (m+1)/2^k \\ 0 & \text{otherwise} \end{cases},$$

then

$$\Pr\{b_n = 2^k\} = \frac{1}{2^k} \quad \text{and} \quad \Pr\{b_n = 0\} = 1 - \frac{1}{2^k}.$$

So, in this example, $b_n \xrightarrow{p} 0$ but $b_n \overset{a.s.}{\nrightarrow} 0$ since the limit of $b_n(\omega)$ does not exist for any ω.

Complete convergence

Definition 3: Let $\{b_n(\omega)\}$ be a sequence of real-valued random variables. If there exists a real-valued random variable $b(\omega)$ such that

$$\sum_{n=1}^{\infty} \Pr\{\omega : |b_n(\omega) - b(\omega)| > \delta\} < \infty,$$

for every $\delta > 0$, then we say that $\{b_n(\omega)\}$ *converges completely* to b.

Proposition 3: Let $\{b_n(\omega)\}$ be a sequence of real-valued r.v. and $b(\omega)$ a real-valued r.v., such that $\{b_n(\omega)\}$ converges completely to b, then $\{b_n(\omega)\}$ converges almost surely to b.

Proof: $b_n \xrightarrow{a.s.} b$ *iff* for every $\delta > 0$,

$$\lim_n \Pr\{\omega : sup_{m \geq n} |b_n(\omega) - b(\omega)| > \delta\} = 0.$$

But,

$$\Pr\{\omega : sup_{m\geq n}|b_m(\omega) - b(\omega)| > \delta\} = \Pr\{\omega : \bigcup_{m\geq n} |b_m(\omega) - b(\omega)| > \delta\}$$

$$\Pr\{\omega : \bigcup_{m\geq n} |b_m(\omega) - b(\omega)| > \delta\} \leq \sum_{m=n}^{\infty} \Pr\{\omega : |b_m(\omega) - b(\omega)| > \delta\},$$

and this sum has limit 0 since $\{b_n(\omega)\}$ converges completely to b.■

Exercise: Show that the reciprocal isn't true.

Hint: Remember that the condition $\lim_n a_n = 0$ is only a necessary condition for the convergence of series $\sum_{n=1}^{\infty} a_n$.

Convergence in rth mean

Definition 4: Let $\{b_n(\omega)\}$ be a sequence of real-valued random variables and r be a positive real number. If there exists a real-valued random variable $b(\omega)$ such that

$$\mathrm{E}[|b_n - b|^r] \to 0 \text{ as } n \to \infty,$$

then we say that $\{b_n(\omega)\}$ *converges in rth mean* to b.

Proposition 4: Let $\{b_n(\omega)\}$ be a sequence of real-valued r.v. and $b(\omega)$ a real-valued r.v., such that $\{b_n(\omega)\}$ converges in rth mean to b, then $\{b_n(\omega)\}$ converges in probability to b.

Proof: Using the generalized Chebyshev inequality,

$$\Pr\{|b_n - b| \geq \varepsilon\} \leq \frac{\mathrm{E}[|b_n - b|^r]}{\varepsilon^r}.$$

■

Convergence in rth mean

But the reciprocal isn't true.

Example 3 (bis): Notice that

$$\mathrm{E}[|b_n - b|^r] = (2^k)^r \frac{1}{2^k} + 0^r (1 - \frac{1}{2^k}) = 2^{k(r-1)},$$

which tends to $\begin{cases} 0 & if \; r < 1 \\ 1 & if \; r = 1 \\ \infty & if \; r > 1 \end{cases}$,

then, $\{b_n\}$ does not converge in rth mean for $r \geq 1$.

Laws of Large Numbers

Definition 5: Let $\{b_n(\omega)\}$ be a sequence of real-valued random variables such that $\mathrm{E}[b_n] = \mu_n < \infty$ and $\bar{b}_n = n^{-1}\sum_{i=1}^{n} b_i$ be the sample mean. If, the sequence $\{\bar{b}_n - \mathrm{E}[\bar{b}_n]\}$ converge in probability to 0, then we say that the $\{b_n(\omega)\}$ satisfy a Weak Laws of Large Numbers (WLLN).

Definition 6: Let $\{b_n(\omega)\}$ be a sequence of real-valued random variables such that $\mathrm{E}[b_n] = \mu_n < \infty$ and $\bar{b}_n = n^{-1}\sum_{i=1}^{n} b_i$ be the sample mean. If, the sequence $\{\bar{b}_n - \mathrm{E}[\bar{b}_n]\}$ converge almost surely to 0, then we say that the $\{b_n(\omega)\}$ satisfy a Strong Laws of Large Numbers (SLLN).

Weak Laws of Large Numbers - 1

Theorem 1: Let $\{b_n(\omega)\}$ be a sequence of real-valued independent r.v. such that:

(a) The r.v. in the sequence are identically distributed.
(b) b_n has finite mean and finite variance.

Then, $\{b_n(\omega)\}$ satisfy the WLLN.

Proof: Since the b_i are i.i.d., then $\mathrm{E}[b_i] = \mu_i = \mu$ and $\mathrm{Var}(b_i) = \sigma_i^2 = \sigma^2$. Therefore,

$$\mathrm{E}[\bar{b}_n] = \mu \quad \text{and} \quad \mathrm{Var}(\bar{b}_n) = \frac{\sigma^2}{n}.$$

Its only rest to use the Chebyshev inequality. ■

Weak Laws of Large Numbers - 2

Theorem 2 (Chebyshev): Let $\{b_n(\omega)\}$ be a sequence of real-valued independent r.v. such that their variance are bounded, i.e., there is a $c > 0$ finite s.t.:

$$\mathrm{Var}(b_n) = \sigma_n^2 \leq c, \text{ for every } n,$$

then $\{b_n(\omega)\}$ satisfy the WLLN.

Proof: Similar to Theorem 1, but noticing that

$$\mathrm{Var}(\bar{b}_n) = \frac{\sum_{i=1}^{n} \sigma_n^2}{n^2} \leq \frac{c}{n}$$

■

Here, we don't required the i.d. assumption.

Weak Laws of Large Numbers - 3

Theorem 3 (Markov): Let $\{b_n(\omega)\}$ be a sequence of real-valued r.v. such that

$$\lim_n \mathrm{Var}(\bar{b}_n) = 0,$$

then $\{b_n(\omega)\}$ satisfy the WLLN.

Proof: Again the Chebyshev inequality.

■

Theorem 4 (Khintchine): Let $\{b_n(\omega)\}$ be a sequence of real-valued independent r.v. such that:

(a) The r.v. in the sequence are identically distributed.
(b) b_n has finite mean.

Then, $\{b_n(\omega)\}$ satisfy the WLLN.

Weak Laws of Large Numbers - 4

Theorem 5 (Gnedenko): Let $\{b_n(\omega)\}$ be a sequence of real-valued r.v. such that $\mathrm{E}[b_i] = \mu_i$. The sequence $\{b_n(\omega)\}$ satisfy the WLLN *iff*

$$\lim_n \mathrm{E}\left[\frac{(\bar{b}_n - \mathrm{E}[\bar{b}_n])^2}{1 + (\bar{b}_n - \mathrm{E}[\bar{b}_n])^2}\right] = 0.$$

An strong result but difficult to apply it.

Strong Laws of Large Numbers - 1

Kolmogorov inequality: Let $\{b_n(\omega)\}$ be a sequence of real-valued independent r.v. such that $\mathrm{E}[b_n] = \mu_n$ and $\mathrm{Var}(b_n) = \sigma_n^2 < \infty$. Then, for every $\delta > 0$,

$$\Pr\left\{\bigcup_{k=1}^{n} \{|S_k - \mathrm{E}[S_k]| \geq \delta V_n\}\right\} \leq \frac{1}{\delta^2},$$

where $S_n = \sum_{k=1}^{n} b_k$ and $V_n^2 = \mathrm{Var}(S_n)$.

Theorem 6 (Kolmogorov): Let $\{b_n(\omega)\}$ be a sequence of real-valued independent r.v. such that $\mathrm{E}[b_n] = \mu_n$, $\mathrm{Var}(b_n) = \sigma_n^2 < \infty$ and

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty,$$

then $\{b_n(\omega)\}$ satisfy the SLLN.

Strong Laws of Large Numbers - 2

Theorem 7 (Khintchine): Let $\{b_n(\omega)\}$ be a sequence of real-valued independent r.v. such that:

(a) The r.v. in the sequence are identically distributed.
(b) b_n has finite mean μ

Then, $\{b_n(\omega)\}$ satisfy the SLLN.

Proof: The proof is slightly long but it use two interesting elements:

- The truncation technique.
- The Borel-Cantelli lemma.

Without lost of generality we can assume that $\mu = 0$.

We construct two sequences of r.v. $\{a_n\}$ and $\{c_n\}$ s.t.:

$$a_n = \begin{cases} 1 & \text{with probability } \rho_n \\ -1 & \text{with probability } 1-\rho_n \end{cases},$$

and the $\{a_n\}$ are independent and independent respect to the $\{b_n\}$.

$$c_n = \begin{cases} b_n & \text{if } |b_n| \leq n \\ na_n & \text{if } |b_n| > n \end{cases}$$

then, $|c_n| = \min(|b_n|, n)$.

The truncation technique.

Moreover, we can find a ρ_n such that $\mathrm{E}[c_n] = 0$:

$$\mathrm{E}[c_n] = \text{ To what?}$$

So,

$$\sum_{n=1}^{\infty} \Pr\{c_n = b_n\} = \sum_{n=1}^{\infty} \Pr\{|b_n| > n\}$$

and

$$\begin{array}{rcl} \Pr\{|b_1| > 1\} & = & \Pr\{1 < |b_1| \leq 2\} + \Pr\{2 < |b_1| \leq 3\} + \cdots \\ \Pr\{|b_2| > 2\} & = & \Pr\{2 < |b_2| \leq 3\} + \Pr\{3 < |b_1| \leq 4\} + \cdots \end{array}$$

but the b_n are i.i.d., then we can write:

$$\begin{array}{rcl} \sum_{n=1}^{\infty} \Pr\{|b_n| > n\} & = & \sum_{n=1}^{\infty} n \Pr\{n < |b| \leq n+1\} \\ & \leq & \sum_{n=1}^{\infty} \int_{n<|x|\leq n+1} |x| dF(x) = \int_{\mathbb{R}} |x| dF(x) < \infty. \end{array}$$

Borel-Cantelli Lemma: Let $\{A_n\}$ be a sequence of random events. If $\sum_{n=1}^{\infty} \Pr\{A_n\} < \infty$, then $\Pr\{\limsup A_n\} = 0$.

By Borel-Cantelli Lemma we have:

$$\Pr\{\bigcap_{m=1}^{\infty}\bigcup_{n\geq m}^{\infty}\{c_n \neq b_n\}\} = 0,$$

or equivalently,

$$\Pr\{\bigcup_{m=1}^{\infty}\bigcap_{n\geq m}^{\infty}\{c_n = b_n\}\} = 1.$$

Then, the truncation c_n is a.s. equal to b_n for a sufficiently large m.

The remaining is to prove that c_n has finite second moment and the Kolmogorov theorem's condition:

$$\sum_{n=1}^{\infty}\frac{\mathrm{E}[c_n^2]}{n^2} < \infty,$$

in order to conclude that $\{c_n\}$ satisfy a SLLN. ■

Convergence in distribution or in law

Definition 7: Let $\{b_n(\omega)\}$ be a sequence of real-valued random variables such that with distribution function $\{F_n\}$. If $F_n(z) \to F(z)$ as $n \to \infty$ for every continuity point z of F and F is the distribution function of a random variable b, then we say that $\{b_n(\omega)\}$ *converge in distribution* to b.

Example 4: Let $\{F_n\}$ be a sequence of distribution functions such that, $F_n(z) = \begin{cases} 0 & \text{if } z < 0 \\ x^n & \text{if } 0 \leq z \leq 1 \\ 1 & \text{if } z \geq 1 \end{cases}$. So, $F_n(x) \to F(x) =?$ for every x, then $b_n \xrightarrow{d} ?$

Example 5: Let $\{F_n\}$ be a sequence of distribution functions such that, $F_n(z) = \begin{cases} 0 & \text{if } z < n \\ 1 & \text{if } z \geq n \end{cases}$, i.e., the r.v. b_n is concentrated on n. Then $F_n(x) \to F(x) = 0$ for every x, then: What we can concluded?

Convergence in distribution - 2

Proposition 5: Let $\{b_n(\omega)\}$ be a sequence of real-valued r.v., $b(\omega)$ a real-valued r.v. and let be f_n and f its densities functions such that $f_n(x) \to f(x)$ as $n \to \infty$ for almost all x. Then, $\{b_n(\omega)\}$ converges in law to b.

Proof: Exercise. ■

Proposition 6: Let $\{b_n(\omega)\}$ be a sequence of real-valued r.v., $b(\omega)$ a real-valued r.v. and let be φ_n and φ its characteristic functions such that $\varphi_n(t) \to \varphi(t)$ as $n \to \infty$ for all t and φ is continuous at $t = 0$. Then, $\{b_n(\omega)\}$ converges in law to b.

Proof: Is more that an exercise. ■

In probability and in law convergence - 1

Proposition 7: Let $\{b_n(\omega)\}$ be a sequence of real-valued r.v. and $b(\omega)$ a real-valued r.v., such that $\{b_n(\omega)\}$ converges in probability to b, then $\{b_n(\omega)\}$ converges in law to b.

Proof: Let

$$\begin{aligned} A_n &= \{b_n \leq z\} = \{b_n \leq z, b \leq y\} \cup \{b_n \leq z, b > y\} \\ &\subset \{b \leq y\} \cup \{b_n \leq z, b > y\}, \end{aligned}$$

then

$$F_n(z) \leq F(y) + \Pr\{b_n \leq z, b > y\}.$$

If $y > z$ then

$$\Pr\{b_n \leq z, b > y\} \leq \Pr\{|b_n - b| > y - z\}.$$

Using that $b_n \xrightarrow{p} b$, we have $\Pr\{|b_n - b| > y - z\} \to 0$

$$\implies F_n(z) - F(y) \leq \Pr\{|b_n - b| > y - z\} \to 0$$

$$\implies \limsup F_n(z) \leq F(y).$$

Analogously, we obtain $\liminf F_n(z) \geq F(x)$ for $x < z$.

Then,

$$F(x) \leq \liminf F_n(z) \leq \limsup F_n(z) \leq F(y)$$

Finally, letting $y \downarrow z$ and $x \uparrow z$, since z is a continuity point of F, we obtain that $F(y) \downarrow F(z)$ and $F(x) \uparrow F(z)$. ■

In probability and in law convergence - 2

The reciprocal isn't true.

Example 6: Let $\{b_n\}$ a sequence of independent and identically distributed r.v. and let b a r.v. such that the joint distribution is give by the following table:

$b : b_n$	0	1	Marginal
0	0	0.5	0.5
1	0.5	0	0.5
Marginal	0.5	0.5	1

Notice that $\Pr\{|b_n - b| > 0.5\} = 1$, then $\{b_n\}$ does not converge in probability to b.

But $b_n \xrightarrow{d} b$ since both has the same distribution function.

In probability and in law convergence - 2

The asymptotic equivalence lemma: Let $\{a_n(\omega)\}$ and $\{b_n(\omega)\}$ be two sequences of real-valued r.v. and $b(\omega)$ a real-valued r.v., such that $\{a_n - b_n(\omega)\}$ converges in probability to 0 and $\{b_n(\omega)\}$ converges in law to b, then $\{a_n(\omega)\}$ converges in law to b.

Proof:

$$\begin{aligned}
\Pr\{a_n \leq x\} &= \Pr\{b_n \leq x + b_n - a_n\} \\
&= \Pr\{b_n \leq x + b_n - a_n, b_n - a_n \leq \varepsilon\} + \\
&\qquad + \Pr\{b_n \leq x + b_n - a_n, b_n - a_n > \varepsilon\} \\
&\leq \Pr\{b_n \leq x + \varepsilon\} + \Pr\{b_n - a_n > \varepsilon\}
\end{aligned},$$

So,

$$\limsup \Pr\{a_n \leq x\} \leq \liminf \Pr\{b_n \leq x + \varepsilon\}.$$

Analogously, we obtain $\liminf \Pr\{a_n \leq x\} \geq \limsup \Pr\{b_n \leq x - \varepsilon\}$.

Its only rest to make $\varepsilon \downarrow 0$ and to use that x is a continuity point. ■

Central Limit Theorem - 1

Definition 8: Let $\{b_n(\omega)\}$ be a sequence of real-valued random variables such that with finite mean and variance. If the sequence $a_n = \frac{\sum_{i=1}^{n} b_i - \mathrm{E}[\sum_{i=1}^{n} b_i]}{\mathrm{Var}(\sum_{i=1}^{n} b_i)^(1/2)}$ converge in law to $\mathcal{N}(0,1)$, then we say that $\{b_n(\omega)\}$ *obeys a central limit theorem*.

Levy-Lindeberg theorem: Let $\{b_n(\omega)\}$ be a sequence of i.i.d. real-valued random variables with finite mean, μ, and finite variance, σ^2. Then $\{b_n(\omega)\}$ *obeys a central limit theorem*.

Proof: Its based on the Taylor expansion of the characteristic functions, $\varphi_{b_k-\mu}(t) = E[\exp(it(b_k - \mu))]$,

$$\varphi_{b_k-\mu}(t) = 1 - \frac{t^2\sigma^2}{2} + o(t^2),$$

where the $o(t^2)$ is the same for all b_k since they are i.d. random variables.

Since the $\{b_n(\omega)\}$ are i.i.d. we have $\mathrm{E}[\sum_{i=1}^n b_i] = \mu$ and $\mathrm{Var}(\sum_{i=1}^n b_i) = \sigma^2$. Then,

$$\begin{aligned} \varphi_{a_n}(t) &= E\left[\exp\left(it\sum_{k=1}^n \frac{b_k-\mu}{\sigma\sqrt{n}}\right)\right] \\ &= \prod_{k=1}^n E\left[\exp\left(it\frac{b_k-\mu}{\sigma\sqrt{n}}\right)\right] \\ &= \varphi_{b_k-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right) \\ &= \left(1 - \frac{t^2\sigma^2}{2\sigma^2 n} + o(t^2)\right)^n \to e^{-\frac{t^2}{2}} \\ &= \left(1 - \frac{t^2}{2} + o(t^2)\right)^n \to e^{-\frac{t^2}{2}} \end{aligned}$$

which is the characteristic function of a $N(0,1)$. ■

Example 7 (De Moivre's theorem): Let $\{b_n\}$ a sequence of independent and identically Bernoulli(p) distributed r.v. then

$$\frac{\sum_{k=1}^n b_k \; - \; np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Liapunov theorem: Let $\{b_n(\omega)\}$ be a sequence of independent real-valued random variables with zero mean and finite variance, σ_n^2. If the $\limsup \frac{E[|b_n|^3]}{s_n^3} = 0$, where $s_n^2 = \sum_{k=1}^n \sigma_k^2$, then

$$\frac{\sum_{k=1}^n b_k}{s_n} \xrightarrow{d} \mathcal{N}(0,1).$$

Lindeberg theorem: Let $\{b_n(\omega)\}$ be a sequence of independent real-valued random variables with finite mean, μ_k, and finite variance, σ_k^2. Let F_k be the distribution function of b_k. If for every $\varepsilon > 0$ the Lindeberg condition

$$\frac{1}{s_n} \sum_{k=1}^{n} \int_{|x-\mu_k| \geq \varepsilon s_n} (x - \mu_k) dF_k(x) \to 0 \text{ as } n \to \infty$$

is satisfy, then

$$\frac{\sum_{k=1}^n b_k - \sum_{k=1}^n \mu_k}{s_n} \xrightarrow{d} \mathcal{N}(0,1).$$

Stochastic orders - 1

Definition 9: The sequence $\{b_n(\omega)\}$ is *at most of order* n^λ *almost surely*, if there exist a $O(1)$ non-stochastic sequence sequence $\{a_n\}$ such that

$$n^{-\lambda} b_n - a_n \xrightarrow{a.s.} 0$$

Notation: $O_{a.s.}(n^\lambda)$

$O_{a.s.}(1)$ is called almost surely bounded

Definition 10: The sequence $\{b_n(\omega)\}$ is of *order smaller than* n^λ *almost surely*, if

$$n^{-\lambda} b_n \xrightarrow{a.s.} 0$$

Notation: $o_{a.s.}(n^\lambda)$

Stochastic orders - 2

Definition 11: The sequence $\{b_n(\omega)\}$ is *at most of order* n^λ *in probability*, if there exist a $O(1)$ non-stochastic sequence sequence $\{a_n\}$ such that

$$n^{-\lambda} b_n - a_n \xrightarrow{p} 0$$

Notation: $O_p(n^\lambda)$

$O_p(1)$ is called bounded in probability

Definition 12: The sequence $\{b_n(\omega)\}$ is of *order smaller than* n^λ *almost surely*, if

$$n^{-\lambda} b_n \xrightarrow{a.s.} 0$$

Notation: $o_p(n^\lambda)$

Stochastic orders - Properties

Proposition 8: Let $\{a_n(\omega)\}$ and $\{b_n(\omega)\}$ be two sequences of real-valued r.v.

(a) If $\{a_n(\omega)\}$ is $O_{a.s.}(n^{\lambda})$ and $\{b_n(\omega)\}$ is $O_{a.s.}(n^{\mu})$, then $\{a_n(\omega)b_n(\omega)\}$ is $O_{a.s.}(n^{\lambda+\mu})$ and $\{a_n(\omega)+b_n(\omega)\}$ is $O_{a.s.}(n^{\kappa})$, where $\kappa = \max(\lambda, \mu)$.

(b) If $\{a_n(\omega)\}$ is $o_{a.s.}(n^{\lambda})$ and $\{b_n(\omega)\}$ is $o_{a.s.}(n^{\mu})$, then $\{a_n(\omega)b_n(\omega)\}$ is $o_{a.s.}(n^{\lambda+\mu})$ and $\{a_n(\omega)+b_n(\omega)\}$ is $o_{a.s.}(n^{\kappa})$, where $\kappa = \max(\lambda, \mu)$.

(c) If $\{a_n(\omega)\}$ is $O_{a.s.}(n^{\lambda})$ and $\{b_n(\omega)\}$ is $o_{a.s.}(n^{\mu})$, then $\{a_n(\omega)b_n(\omega)\}$ is $o_{a.s.}(n^{\lambda+\mu})$ and $\{a_n(\omega)+b_n(\omega)\}$ is $O_{a.s.}(n^{\kappa})$, where $\kappa = \max(\lambda, \mu)$.

Proposition 8 holds for *in probability* orders

Proposition 9: Let $\{b_n(\omega)\}$ be a sequence of real-valued r.v. and b be a r.v., such that $b_n \xrightarrow{d} b$, then $b_n = O_p(1)$.

Asymptotic properties of the least squares estimator

Consistency:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta+\varepsilon) \\ &= \beta_0 + (X'X)^{-1}X'\varepsilon \\ &= \beta_0 + \left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}\end{aligned}$$

$\boxed{\left(\frac{X'X}{n}\right)^{-1}}$: By assumption $\lim_{n\to\infty}\left(\frac{X'X}{n}\right) = Q_X \Rightarrow \lim_{n\to\infty}\left(\frac{X'X}{n}\right)^{-1} = Q_X^{-1}$, since the inverse of a nonsingular matrix is a continuous function of the elements of the matrix.

$\boxed{\frac{X'\varepsilon}{n}}$:

$$\frac{X'\varepsilon}{n} = \frac{1}{n}\sum_{t=1}^{n} x_t\varepsilon_t.$$

Each $x_t\varepsilon_t$ has expectation zero, so $\mathrm{E}\left(\frac{X'\varepsilon}{n}\right) = 0.$

The variance of each term is $\mathrm{Var}\left(x_t\epsilon_t\right) = x_t x_t'\sigma^2.$

As long as these are finite, and given a technical condition, the Kolmogorov SLLN applies, so

$$\frac{1}{n}\sum_{t=1}^{n} x_t\varepsilon_t \stackrel{a.s.}{\to} 0.$$

This implies that $\hat{\beta} \stackrel{a.s.}{\to} \beta_0.$

Asymptotic properties of the LSE - 2

Normality:

$$\begin{aligned}\hat{\beta} &= \beta_0 + (X'X)^{-1}X'\varepsilon \\ \hat{\beta} - \beta_0 &= (X'X)^{-1}X'\varepsilon \\ \sqrt{n}\left(\hat{\beta} - \beta_0\right) &= \left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{\sqrt{n}}\end{aligned}$$

- $\left(\frac{X'X}{n}\right)^{-1} \to Q_X^{-1}$.

- Considering $\frac{X'\varepsilon}{\sqrt{n}}$, the limit of the variance is

$$\begin{aligned}\lim_{n\to\infty} \mathrm{Var}\left(\frac{X'\varepsilon}{\sqrt{n}}\right) &= \lim_{n\to\infty} \mathrm{E}\left(\frac{X'\epsilon\epsilon'X}{n}\right) \\ &= \sigma_0^2 Q_X\end{aligned}$$

- We assume one (for instance, the Lindeberg CLT) holds, so

$$\frac{X'\varepsilon}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\left(0, \sigma_0^2 Q_X\right)$$

  Therefore,

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_0^2 Q_X^{-1}\right)$$

- In summary, the OLS estimator is normally distributed in small and large samples if ε is normally distributed. If ε is not normally distributed, $\hat{\beta}$ is asymptotically normally distributed when a CLT can be applied.

Asymptotic properties of the LSE - 3

Asymptotic efficiency

The least squares objective function is

$$s(\beta) = \sum_{t=1}^{n} \left(y_t - x_t'\beta\right)^2$$

Supposing that ε is normally distributed, the model is

$$y = X\beta_0 + \varepsilon,$$

with

$$\varepsilon \sim \mathcal{N}(0, \sigma_0^2 I_n),$$

so

$$f(\varepsilon) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right)$$

The joint density for y can be constructed using a change of variables: $\varepsilon = y - X\beta$, so $\frac{\partial \varepsilon}{\partial y'} = I_n$ and $|\frac{\partial \varepsilon}{\partial y'}| = 1$, then

$$f(y) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - x_t'\beta)^2}{2\sigma^2}\right).$$

Taking logs,

$$\ln L(\beta, \sigma) = -n \ln \sqrt{2\pi} - n \ln \sigma - \sum_{t=1}^{n} \frac{(y_t - x_t'\beta)^2}{2\sigma^2}.$$

It's clear that the objective function for the MLE of β_0 are the same as the objective function for OLS (up to multiplication by a constant), so the ML estimator and OLS share their properties.

In particular, under the classical assumptions with normality, the OLS estimator $\hat{\beta}$ *is asymptotically efficient.*